LA-UR-13-28429
Approved for public release; distribution is unlimited.

Title:	Final Report on C8 Enhanced Surveillance Campaign Milestone L2 4654: Surveillance Metrics Assessments on Selected Weapons
Author(s):	Ticknor, Lawrence O. Huzurbazar, Aparna V.
Intended for:	Surveillance Metrics Assessment, 2013-11-05 (Washington, District Of Columbia, United States)
Issued:	2013-11-18 (rev.1)

Final Report on C8 Enhanced Surveillance Campaign Milestone L2 4654:

Surveillance Metrics Assessments on Selected Weapons

Lawrence Ticknor, CCS-6
Statistical Sciences Group
Enhanced Surveillance Campaign

Los Alamos National Laboratory

Dr. Aparna Huzurbazar
Project Lead, Systems MTE,

U.S. DEPARTMENT OF ENERGY

NNSA National Nuclear Security Administration DEFENSEPROGRAMS

LA-UR-13-28429

Introduction and Outline

- Brief Overview of C8 Enhanced Surveillance
- Assessment Team Qualifications
- ES Surveillance Metrics Milestone
- Overview of Approach 1 and Approach 2
- Dashboards
- Differences and Similarities – Data and Models
- Opportunities – Sensitivity Analyses (“What if”)
- Short Comparison Summary

Enhanced Surveillance – Mission & Objectives

From Diagnostics to Lifetimes

Campaign 8 Enhanced Surveillance

The Enhanced Surveillance Subprogram contributes to weapon safety, performance and reliability by providing tools needed to predict or detect the precursors of age-related defects and to provide accurate engineering estimates of component or system lifetimes.*

Primary Objectives

- **Identify stockpile aging behavior** using available diagnostic tools
- **Develop new cost effective capabilities tools/diagnostics and new methods (Diagnostics & CME)**
  - Includes development and use of sensor technology for in-situ aging studies – surveillance of the future
- **Understand aging behaviors**
  - determine aging mechanisms
- **Provide improved predictive capabilities (materials aging models**
  - Inform stockpile decisions on Annual Assessments, SFIs and LEPs
- **Perform lifetime assessments** in support of refurbishment schedules
  - direct line to physics performance using latest ASC codes

From the Identification of Defects

To Lifetime Estimates

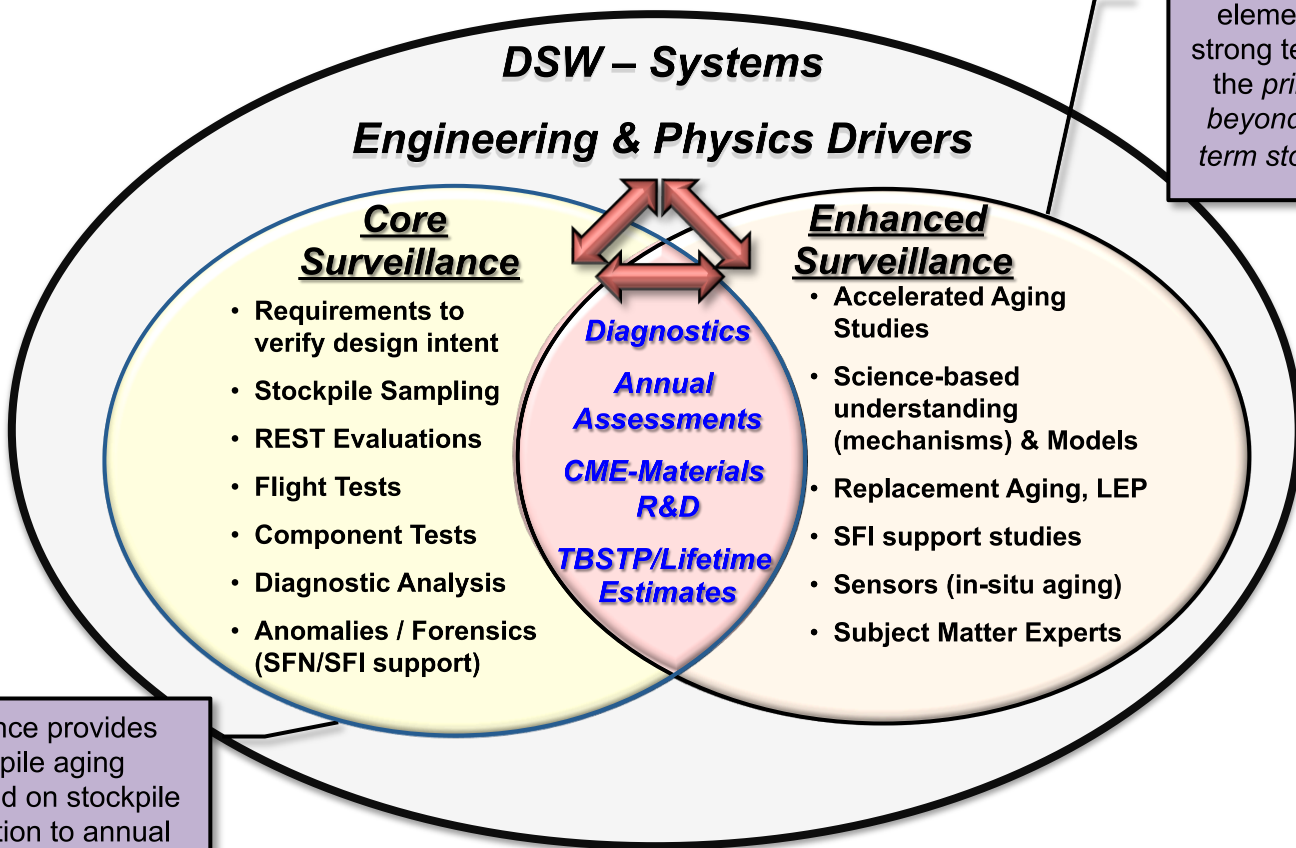
… AGING, AGING, AGING

*FY10 Engineering Campaign Program Implementation Plan

Linkages between Core and Enhanced

Two different, but complementary programs

ES contributes to critical elements of DSW through strong technical diversity, with the *primary goal* of *looking beyond the horizon for long-term stockpile aging behavior*



Core Surveillance provides critical stockpile aging information based on stockpile data for application to annual assessments

Core and Enhanced Surveillance Programs collaborate in a number of key areas

REST - Retrofit Evaluation System Test ; SFN – Significant Finding Notification; SFI – Significant Finding Investigation; CME –Component Material Evaluation; TBSTP - Technical Basis for Stockpile Transformation Planning

Assessment Team Qualifications

- Lawrence Ticknor: 20+ years at LANL in the Statistical Sciences Group.
  - Lean Six Sigma Master Black Belt
  - In addition to weapons, has worked on Safeguards (measurements, shipments, instrumentation certification), sampling plans, Biowatch, bio-weapons detection, genomics)
- Geralyn Hemphill: 27 years experience in NW complex, including Rocky Flats.
  - Lean Six Sigma Master Black Belt
  - Has worked for ES, Core Surveillance, B61, W76/78.
- Aparna Huzurbazar, PhD: 6 years at LANL+13 years University of New Mexico and RAND Corporation.
  - Lean Six Sigma Master Black Belt
  - PL Systems MTE, ES. Also works for C5 and B61-LEP
  - Author of: *Flowgraph Models for Multistate Time-to-Event Data* (Wiley, 2005)

C8 Milestone L2 4654
Surveillance metrics assessments on selected weapons

Grading Criteria

- Enhanced Surveillance will apply Approach-1 to one weapon system and provide a comparison of the two approaches to that system.
- ES will examine opportunities to apply and integrate CS/ES data and/or models to quantify LANL's ability to assess understanding of long-term material and component behaviors.

Exit Criteria

- ES will compile and communicate findings of the assessment in a report with internal LANL distribution and submittal to the NNSA ES subprogram FPM (NA-124).
- Final Report on C8 Enhanced Surveillance Campaign Milestone L2 4654: Surveillance Metrics Assessments on Selected Weapons LA-UR-13-28095

Sources of Information

Worked closely with Core and ES at LANL and LLNL

- Geralyn Hemphill (LANL) provided information on weapons systems, expert knowledge and context
- Aparna Huzurbazar (LANL) provided information on weapon systems and overall project context and direction.
- Bill Mclean(LLNL)provide lots of time and information on LLNL Approach 1 – sharing time, spreadsheets, ideas, papers, etc. and looking at my work on Approach 1
- Mike Hamada provided lots of information on the development of Approach 2 – context, history, papers, thoughts, strategies, etc.
- Michael Peters (LANL – W88 Data), Jeff Abes (LANL – Implementing Approach 2) provided data, spreadsheets, information, context, results, dashboards, etc. for Approach 2 and specifically the W88 system.
- Tom Zocco (LANL) and Charles Hill (LANL) provided context on their programs and help understanding history and surveillance in general
- Our focus was the NEP and applications of metrics to

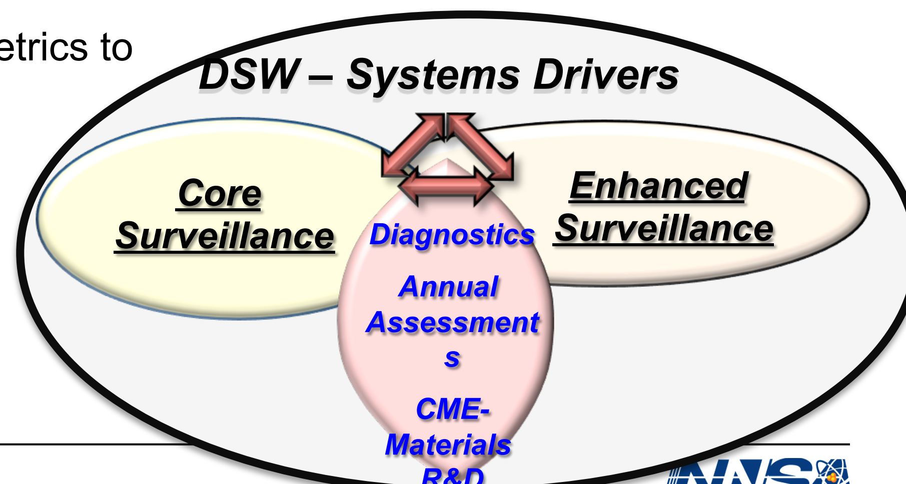
NEP so we did not seek support from SNL

[Any omissions, errors, controversies, etc.

are attributed to Larry Ticknor]

[Providing Information or Data does not mean

agreement with the conclusions]



Metrics Respond to Dr. Greenaugh/Kusnezov Memo of August 10, 2009

- 1. *A more quantitative and/or qualitative metric(s) describing the results of realized and non-realized surveillance activities on our confidence in reporting reliability and assessing the stockpile.* As discussions ensue for resource allocations in future years for the Surveillance Enterprise, we must enhance our ability to *quantitatively and/or qualitatively communicate* the results of realized or not realized surveillance activities. This study will provide recommended metrics and associated implementation strategy, endorsed by the three laboratories, to accomplish *clear and effective communication* to external stakeholders, including the Department of Defense, Executive Branch offices, and Congressional bodies.
- Both approaches agree that their metrics are to measure *"the health of the surveillance program [and] not the weapon."*
- *Predictive Assessment* is a part of both approaches and the Dr. Greenaugh memo

Approach 1

Approach 1 [LLNL – William (Bill) McLean]

- “designed to aid managers“
- “small lots of expensive but irreplaceable assets”
- “simple calculations that establish the relative influence of data acquisition from stockpile surveillance and the quality of physical/ chemical models of that data. … it allows one to exploit multiple data and knowledge streams in order to reduce the need for statistically significant annual sampling of a small nuclear stockpile”
- ”easily visualize”
- “the list [of scores] developed and prioritized in this way is, in itself, a program execution plan for the weapon project managers and other campaigns to follow”

Approach 1 – Details

Wxx Subsystem Stockpile Knowledge Metric Score														Surveillance Execution Score			
	Detect Defects Score = SCF x DCS								Model Score								
	Surveillance Confidence Factor				Data Currency Score												
Subsystem	Population Size (N)	Samples (n)	R = 1 - (Undetected Defect Fraction)	(SCF)	Years until confidence goes to zero	Years since last surveillance (YSLS)	(DCS)	(DDS)	Current unit age	Design Lifetimen (years)	Confidence in Aging Model	(MS)	(SKM)	FC = Fraction of program of record completed in last 4 years	RRT = Requirements Review Toggle	(SES)	Final Score = SKM x SES
NNC	500	35	0.90	**0.98**	5	1.0	**1.00**	**0.98**	23	30	L	**0.55**	**0.76**	0.93	1.00	**0.93**	**0.71**
HE	500	25	0.90	**0.93**	5	2.0	**0.75**	**0.70**	23	30	M	**0.65**	**0.67**	0.83	1.00	**0.83**	**0.56**
PIT	500	20	0.90	**0.88**	5	3.0	**0.50**	**0.44**	23	30	H	**0.80**	**0.62**	0.89	1.00	**0.89**	**0.55**
CSA	500	9	0.90	**0.62**	5	1.0	**1.00**	**0.62**	23	30	L	**0.55**	**0.58**	0.50	1.00	**0.50**	**0.29**

$$\text{Final Score} = \text{SKM} \times \text{SES} = \left[\frac{(\text{SCF} \times \text{DCS}) + \text{MS}}{2}\right] \times \text{FC} \times \text{RRT}$$

Visual management tool

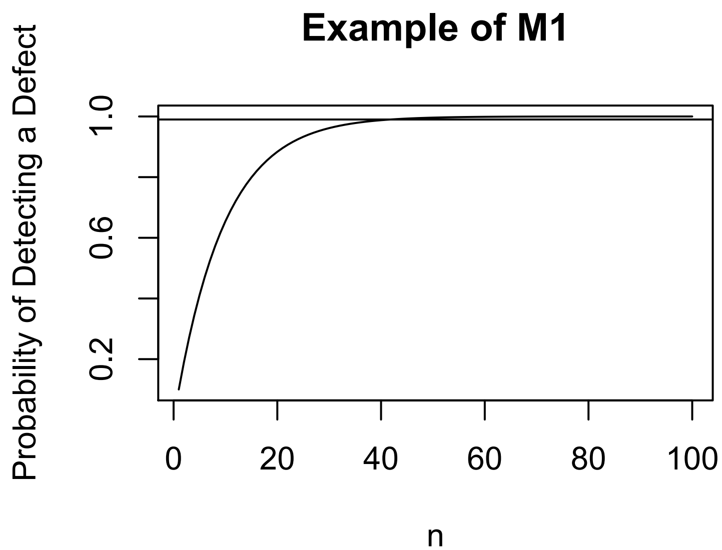
Approach 2

Approach 2 – LANL/SNL [Statisticians /Mathematicians]

- “Surveillance metrics” “assess the adequacy of data/information collected”:
  1. “To detect static and latent defects”
  2. “To assess margins [present] and future”
  3. “To validate predictions”
- “measure how well we know what we know”
- “the surveillance metrics are statistically based assessments”
- “surveillance metrics are probabilities/confidences ranging from 0 to 1, where 0 is woefully inadequate and 1 is extremely adequate relative to an accepted level of risk”
- “If a metric is low, more data needs to be collected to improve the metric.”

Approach 2 – Details of M1

M1 – Probability of Detecting a Static (M1a) or Latent Defect (M1b)

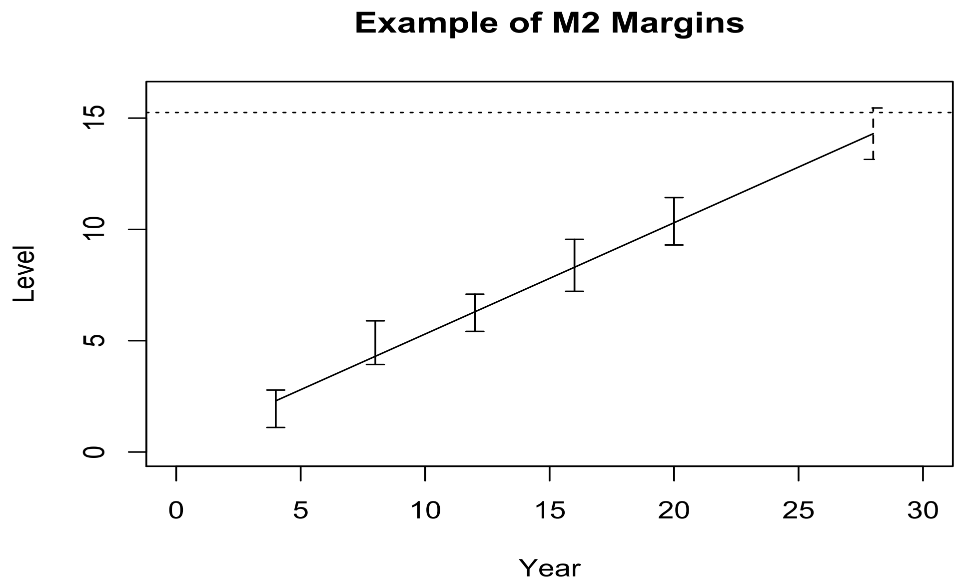


M1a = "Surveillance Confidence Factor" of Approach 1

Approach 2 – Details of M2

M2 – Assess Present and Future Margins

Confidence that data are different than some specification



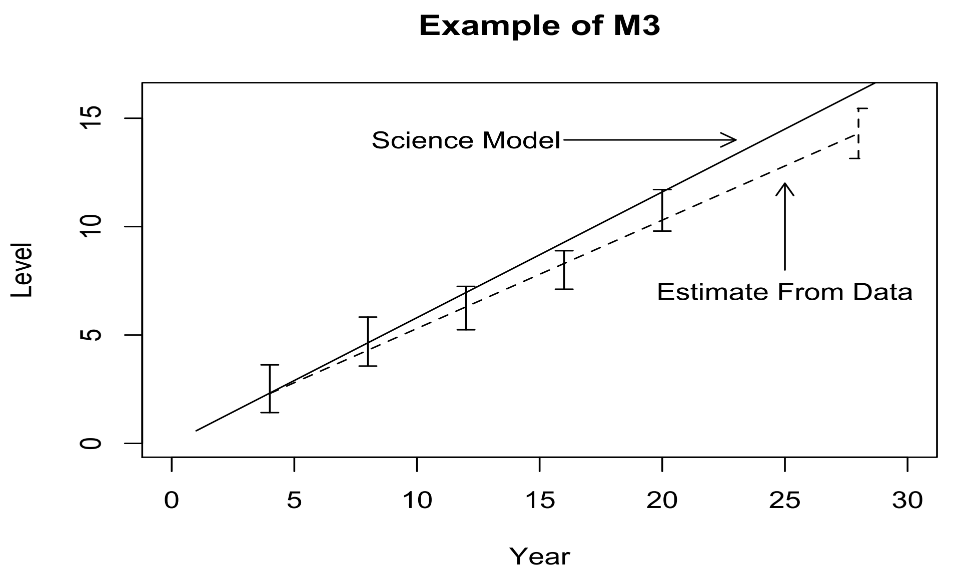
Failure Specifications Often Do Not Exist

Use Manufacturing Tolerances or Other Specifications

Approach 2 – Details of M3

M3 – Validate Predictions – Magnitude of Difference Between Data and Prediction

Probability of Detecting a Change in Mean (M3a) or Standard Deviation (M3b) From a Predicted Value



Probabilities Depend on Sample Size "n" and the Magnitude of Deviation to Detect. Large Deviations are More Likely to Be Detected.

Approach 2 – Details

Approach 2 tests give "confidence" of determining a statistical difference

- High Metric Values Do Not Imply High Confidence in Component
- High Metric Values Imply High Confidence in Decision
- Low Metric Values Imply Not Enough Data for Statistical Decision

- Low Metric Values Suggest May Need To:
  a) Collect More Data
  b) Look For Other Data Sources
  c) Lower Uncertainty in Failure Specification (M2)
  d) Lower Uncertainty in Prediction
  e) Decrease Errors in Data
  f) Others…

Dashboards and Comparing Results

Approaches have different emphasis

"Dashboard" or "Rolled Up" Values Depend on Weights to Combine Individual Test Metrics

Example Subsystem "Dashboard" Values		
	High	Low
Approach 1 – High Risk	0.48	0.00
Approach 1- Average	0.60	0.11
Approach 2	0.90	0.07

Cannot compare across approaches, only compare within approaches

High Values Suggest Component is Acceptable
Low Values Suggest Component Needs Attention

Only Weights Were Changed to Go From High to Low – Data Never Changes

- One Test Had 0 Samples
- Test M1a or Surveillance Confidence Factor was 0.0 for Both Approaches

Cannot Compare Results Without Comparable Weights – Approaches Have Very Different Metrics => No Comparable Weights

Highly variable results depending on weighting

Models and Model Quality – Approach 1

Approach 1 differs from Approach 2:
Based on the Premise:

Models Can Substitute For Data

High Quality Models are Defined by:

- Based on Scientific Principles
- Based on Accelerated Aging Studies
- “Correctly” Predict Data
- “Reproduce” Stockpile Returns
- Well Developed Uncertainty Quantification Process
- Capable of Predicting 20 Years Into the Future

Low Quality models are

- Statistical fits to noisy data and
- No scientific basis to predict future behavior

Difficult to Quantify “Correctly”, “Reproduce”, “Well Developed”

All Models Can Predict 20 Years into the Future – Need to Understand Accuracy of the Prediction

Wxx Subsystem Stockpile Knowledge Metric Score															Surveillance Execution Score			
	Detect Defects Score = SCF x DCS								Model Score									
	Surveillance Confidence Factor				Data Currency Score													
Subsystem	Population Size (N)	Samples (n)	R = 1 - (Undetected Defect Fraction)	(SCF)	Years until confidence goes to zero	Years since last surveillance (YSLS)	(DCS)	(DDS)	Current unit age	Design Lifetimen (years)	Confidence in Aging Model	(MS)	(SKM)	FC = Fraction of program of record completed in last 4 years	RRT = Requirements Review Toggle	(SES)	Final Score = SKM x SES
NNC	500	35	0.90	0.98	5	1.0	1.00	0.98	23	30	L	0.55	0.76	0.93	1.00	0.93	0.71
HE	500	25	0.90	0.93	5	2.0	0.75	0.70	23	30	M	0.65	0.67	0.83	1.00	0.83	0.56
PIT	500	20	0.90	0.88	5	3.0	0.50	0.44	23	30	H	0.80	0.62	0.89	1.00	0.89	0.55
CSA	500	9	0.90	0.62	5	1.0	1.00	0.62	23	30	L	0.55	0.58	0.50	1.00	0.50	0.29

Rating Models is Difficult – Requires Expert Knowledge – Agreement Between Experts May Be a Problem

Models and Model Quality (Approach 1)

Rating Models for Approach 1: An Example

Models Exist for Reaction Rates of Au Wires in Pb-Sn-In Solder

Rate Depends on Age and Temperature

Is This Model High Quality?

Modeler Says “High Quality”

- Developed From Accelerated Aging Studies
- Predicts Data Well and Reproduces Stockpile Returns

User Says “Not Useful”

- Model Depends on Temperature – No Reliable Temperature Data

Manager Says “Useful”

- Assume Maximum Likely Temperature
- Compare Results to Failure Specification
- If Prediction From Max. Temperature is “Below” Failure Spec. Model is Useful

Models and Model Quality (Approach 2)

Approach 2 use of Models

Models Used in M2 (Margins) and M3 (Predictions)

- M2 – Compares Predictions to Failure Specifications
- M3 – Determines the Probability of Detecting Differences in Means and Standard Deviations Between Prediction and Data

All Model Errors Not Incorporated
Models Not Rated as in Approach 1 –
Model Quality Not Incorporated into Approach 2 (beyond prediction uncertainty)
Approach 2 Concerned With Data to Compare to Models

Models Tied to Goal of Predicting Reliability and Assessing Stockpile

Do Need Models

Do Need to Know How Good Are Those Models

Data Quantity

Approach 1 and 2 Agree – “Risks in Not Collecting Data”

Approach 1

- Looks At Total Number of Samples Collected
- Models Can Substitute for Data
- Requires Periodic Collection of Data (Timing is Important)
- Number of Samples Satisfies “Program of Record”
- No Metric if Number of Samples Too Few or Too Many
- Benefit to “Surveillance Confidence Factor” of More Samples Diminishes Quickly

Approach 2

- Requires Data for all Metrics
- M1b, M3a, M3b based on “Current” Data
- Considers If Enough Data to Make Decision – Not Directly if Too Much
- “Allows” Current Data to be Clustered in Time (ex. All in 1 year of 4)
- Requires More “Current” Data Than Approach 1 (Generally)

Data Quality

Data Quality Includes:

Measurement Errors

- Random Errors
- Systematic Error
- Misclassification Errors

Representative of Failure Rates of Modes of Interest

Representative of Factor of Interest

Uncertainties in Failure Specifications

Randomness of Samples

Comparability to Other Data

Metrics, In General, Do Not Consider Data Quality

Model / Data Connections (Approach 1)

Models Might Allow Less Frequent Sampling

Models Should Lower Probability of Unknowns – “Failure Cliffs”

Validated Accelerated Aging Tests Might Lower Probability of Unknowns

- Pass/Fail Tests (Often) Do Not

Models Need to Predict With Low Uncertainty

Implementing Metrics – Observations and Recommendations

After First Time – Metrics Not as Time Consuming

- First Time - Understanding and Gathering Information Time Consuming
- After First Time –Need to Incorporate Changes in Data and Models
- Always – Time Consuming to Look at Data

Optimally – Data in Databases and Metrics Connected to Databases

Weights – Determine How Combine Information

- Need Documentation
- Need to Be Updated When More Information / Better Techniques are Obtained

Need “Acceptable Level Of Risk”

- Is it “90% Confidence that 90% of the Population” is Good or
- Is it “99% Confidence that 95% of the Population” is Good

- M3 Written to Look at Multiples of Standard Deviation – What is of “Practical Importance”

Opportunities For Improving Metrics – Sensitivity to Data Quantity

Determine How Metric Values Change with:

- Additional Samples for One or More Years
- Fewer Samples for One or More Years

Are Metric Value Differences of Practical Importance?

Approach 1: Limited Opportunities

- Only Data Quantity Entry is For Surveillance Confidence Factor
- After 30 Samples, the Change in Value With Added Samples is Minimal

Approach 2: Many Opportunities

- All Metrics Based on Data
- Some Metrics More Sensitive to Additional or Fewer Samples

Opportunities For Improving Metrics – Sensitivity to Timing of Data Collection

Determine How Metric Values Change with:

- Not Collecting Data Every Year –
  - Collect 2 Samples One Year and 0 the Next or …
- Are Metric Value Differences of Practical Importance?
- Approach 1: Includes Factors Looking at Data Collection Frequency and Loss of Confidence Due to Not Collecting Data for 1 or More Years
  - Major Information to Change Would be How Confidence Decreases With Not Collecting Data
- Approach 2: Many Opportunities
  - Metrics Based on “Recent” Data Would be Most Affected (M1b, M3)
  - The Number of Years Considered “Recent” and the Number of Years Data Is Not Collected Both Contribute to the Result
- *Some Measurements Must be Made Regularly to Ensure Capability So a Staggered Schedule of Measurements May Not be Feasible*

Opportunities For Improving Metrics – Sensitivity to Measurement Errors

Determine How Metric Values Change with:

- Different Values of Measurement Errors –
  - Random Errors, Systematic Errors, Misclassification Errors
  - Neither Method Incorporates These Errors into the Metrics (In General)
  - **Suggestion**: Improve Metrics to Include These Errors Within Metrics and Set to 0 if Needed

- Approach 1: Limited Opportunities
  - Add Misclassification Error for Surveillance Confidence Factor
  - Include Measurement Errors in Rating Models

- Approach 2: Many Opportunities
  - Sensitivity to Maximum Expected Errors -- Record How Metrics Change
  - Increases Confidence in Reported Results
  - Errors That Have Greatest Effect on Confidence Could be Determined

Opportunities For Improving Metrics – Sensitivity to Models and Model Quality

Campaign 8 Enhanced Surveillance

Determine How Metric Values Change with:

Models:

Approach 1: Metric Values Depend on Existence of Model - Easy to See Metric Changes
Approach 2: Metric Values Depend on Predictions – No Model, No Metric Value, Problem?

Model Quality:

How Assess Models' Usefulness/Quality/Importance

a) Model Prediction Uncertainty
b) Data Quality
c) Data Quantity
d) Assumptions
e) Representativeness
f) Inputs Availability and Quality
g) Etc.

To Assess When Models Will Improve Confidence/Understanding of Stockpile

Sensitivity of Metrics to:

a) New Model
b) Improved Model
c) More Data
d) Data With Lower Measurement Uncertainties,
e) Failure Specifications With Lower Uncertainty

Tradeoffs -- Cost of Information versus Improvements in Knowledge of "Reliability"

Approach 1: Easy to See Effect of Changing Model Ratings on Metric Values: Changes in Data or Models Hidden in Model Ratings
Approach 2: Possibilities Exist for Data Changes – Model Changes Not as Obvious Since Model Quality Not Directly Considered – But Many Possibilities Exist
[High Metrics Values in Every Metric May Not Be Needed or Cost Effective]

Grand Challenge – Metrics To Assess When Models Will Improve Confidence/ Understanding in Surveillance Activities

Cautions
To Be Useful Metrics Need to Be More Than an "Exercise"

Surveillance Data Must Be Studied

- Resist Automating and Computer Generating Metrics
- Raw Data Up Through Metrics Must Be Critically Viewed

Metrics Require Knowledgeable Experts

- Surveillance Experts
- Weapons Systems Experts
- Measurement Experts
- Statisticians (Number Experts)

"Dashboard" Values Depend on Weighting Schemes

- Only Believe the Single Value If the Weighting Scheme is Well Known and You Agree With the Weights
- Spend Time Understanding the Individual Component Test Results

High Values in Every Metric May Not Be Needed or Cost Effective

Quick General Comparison Summary

For Metric Values	Approach 1	Approach 2
Data		
Counts of Data	Yes	Yes
Uses Data Values	Yes*	Yes
Rates Data Quality	No	No
Models		
Uses Model Values	No	Yes
Rates Model Quality	Yes	No
Failure Criteria (Failure Specifications)		
Uses Criteria	No	Yes
Rates Criteria Quality	No	No
Weights for Dashboards		
Need Better Documentation	Yes	Yes

*** 11/7/2013 – Bill McLean changed Approach 1 to include a factor for data use.**

Both Approaches are Being Used and Benefitting Surveillance Programs

Questions?

Thanks for your Attention.

References

Bierbaum, Rene; Hamada, Michael; Robertson, Alix (2011). *Surveillance Metrics Sensitivity Study*, SAND2011-6763.

R. Bierbaum, K. Diegert, M.S. Hamada, A. Huzurbazar and A. Robertson (2012a), "Surveillance metrics", *Los Alamos National Lab Technical Report* LA-UR-12-22697.

R. Bierbaum, K. Diegert, M.S. Hamada, A. Huzurbazar and A. Robertson (2012b), "Using statistical methods to assess a surveillance program". *Los Alamos National Lab Technical Report* LA-UR-12-22724. Currently submitted to *Quality Engineering* (revised LA-UR-13-25760).

Greenaugh, Kevin G. and Kusnezov, Dimitri (2009), "National Nuclear Security Administration Comprehensive Surveillance Enterprise Study," DOE Memorandum, 10 August 2009.

M.S. Hamada, G.M. Hemphill, A.M. Peters, J.I. Abes (2013). "Incorporating science models and experimental data into the evaluation of the M2 surveillance metric", *Los Alamos National Lab Technical Report* , LA-UR-13-24582.

M. Hamada (2012). "Metric M2b/M2c for nonlinear regression model", *Los Alamos National Lab Technical Report* LA-UR-12-26012 .

K. Henson, K. Sentz, M. Hamada (2012). Weighting and Bayes nets for rollup of surveillance metrics, *Los Alamos National Lab Technical Report* LA-UR-12-20962.

McLean, William (2011a): William McLean (LLNL), Alix Robertson (LLNL), Mike Hamada (LANL), Charles Hills (LANL), Aparna Huzurbazar (LANL), Earl Whitney (LANL), Jeff Abes (LANL), Rene Bierbaum (SNL), Kathleen Diegert (SNL/NNSA). *FY11 Report of the Surveillance Metrics Team: Status of Metric Development and Implementation (U)*

McLean, William (2011b): "Managing the Surveillance Dilemma (U)," (2011) Lawrence Livermore National Laboratory Document LLNL-AR-506251.

McLean, William (2012), "A Simple Metric to Assess the State-of-Surveillance," CODT-2012-0459 (March 30, 2012)

B.P. Weaver, M.S. Hamada (2012). "An assessment of the sensitivity of the surveillance metrics", *Los Alamos National Lab Technical Report* LA-UR-12-24899

Extra Slides

Improved Understanding of both approaches

Campaign 8 Enhanced Surveillance

Approach 2 is intended to gauge the adequacy of core surveillance. Are we collecting enough data to:

M1: Detect static or aging defects in the stockpile

M2: Determine if a trend currently or in the future is "close" to a specification

M3: Identify deviations (in mean or variability) from a model

Approach 1 (as used by LLNL) includes M1, a measure of completed core surveillance requirements , and incorporates a subjective assessment of knowledge obtained from science-based models and accelerated aging studies used for projecting the future state of the stockpile. A metric is computed based on the average impacts within a subsystem.

What is missing are metrics that would identify where better models are needed for predicting long-term component and material behavior and what studies need to be done in order to better understand critical limits related to component and material performance.

Approach 2 Weighting Example

Not all Approach 2 Metrics Would Have Same Weighting

If M1b is high (high probability of detecting a defect in the population using just the recent data – then M1a score (which uses recent and old data) would also be high and the result is redundant

If M1b, M2b, or M2c is low and there is a high quality model (as per Approach 1 definition) then having the model means the lack of data for M1b is not as important.

If M2c is high, then M3 values are not as important

What is Required for Predictive Ability

Campaign 8 Enhanced Surveillance

If Confidence in Reporting Reliability and the Stockpile Requires Predictive Ability -- What is Required?

- Data
  - Existing State of Components
  - Information for Models
  - Changes in Components
- Models
  - Predict Future State of Components
  - Predict Future Margins
- Failure Specifications
  - Margins Needed to Judge When More Information is Needed and What Type of Information

Opportunities Exist For Improving Surveillance Metrics

Metrics That Measure Effect of Additional or Fewer Samples

- Could be Done with “What-if” Scenarios as Both Approaches Are Doing Now

Data With High Variability - Lowers Metric Values

- Collect More Data or Less Variable Data
- Would Like to Know Effect of More Data and Less Variable Data
- Suggests What, If Any, Changes Might Help

Data With a Bias – If Known Bias can Correct.

- If Bias Changes Through Time -- Harder to Correct
- Affects How Well Know Margin
- Need to Ask if Bias Exists
- If Unknown – Find Maximum Bias That Has No Practical Effect on Results

Opportunities Exist For Improving Surveillance Metrics

Increased Recognition of Measurement Errors

- Random Errors, Systematic Errors, Misclassification Errors
- Improve Metrics to Have “Place Holders” For These

1. Shows what information would like to use
2. Allows “what-if” exercise – Look at Minimum and Maximum Expected Errors and Record How Results Change
3. Increases Confidence in Reported Results
4. Suggests What Errors Might Be Most Important

Opportunities Exist For Improving Surveillance Metrics

Models

How Assess Models as “Low”, “Medium”, or “High” Usefulness/Quality/Importance

a) Model Prediction Uncertainty
b) Data Quality
c) Data Quantity
d) Assumptions
e) Representativeness
f) Inputs Availability and Quality
g) Etc.

Metric of When Models Will Improve Confidence/Understanding of Stockpile

Low Metric Values Suggest More Information is Needed –

Is A Measure Possible that Ranks a Model

a) Versus More Data,
b) Versus Lower Measurement Uncertainties,
c) Versus Better Failure Specifications (Margins)

Tradeoffs -- Cost of Information versus Improvements in Knowledge of “Reliability”

[High Metrics Values in Every Metric May Not Be Needed or Cost Effective]